Stock Market Analysis and Prediction using Data Science

Shani Kumar
Bellevue University
Bellevue, Nebraska 68005, USA
shkumar@my365.bellevue.edu

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored.
For all other uses, contact the Owner/Author.
Copyright is held by the owner/author.

Abstract
The Stock Market is a complex world in itself, and the stock price movement is hard to predict. Accurate predictions on how stocks are going to perform remain a crucial issue in finance. Participants in trading are constantly looking for techniques that can better predict stock price movements. Technical and Fundamental indicators based on the past performance of stocks can help us understand how the stocks are going to perform. However many factors can impact market sentiments like political news, social media buzz, and the latest developments in the company are a few of the examples which can change the stock trends. A massive amount of data gets generated daily hence predicting stock movement in a limited time is very challenging. We can use the Data Science processes to create machine learning models that can predict the stock price movement better than standard practices used in traditional strategies.

Author Keywords
Stock Market; Prediction; Social Media - Twitter, Facebook, Sentimental Analysis, Natural Language Processing (NLP), Artificial Neural Networks (ANN), Support Vector Regression (SVR), Back Propagation Neural Networks (BP-Neural Nets), Support Vector Machines (SVM), Independent component analysis (ICA).

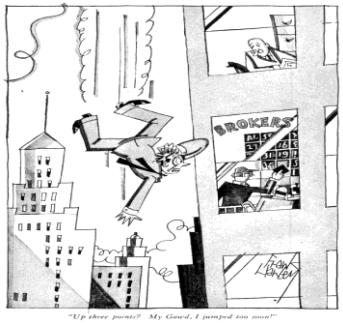

Figure 1: Good preparation and market knowledge required for investments otherwise the same can happen to any investor.

Photo CC-BY Eduardo Bolinches on Twitter.

ACM Classification Keywords
I.5.1 Pattern Recognition: Models; I.2.7 Natural Language Processing: Text Analysis.

Introduction

Almost unlimited uncertainties make it very difficult to predict price movement in the Stock Market. We can see a lot of development in this space. A variety of factors indicate the probable future direction of a stock. The most commonly used variables are stock's market capitalization, earnings per share (EPS), price to earnings ratio (P/E ratio), debt to equity ratio, return on equity ratio, profit margin, return on assets, dividend yield, etc. We can get these details from the online portal easily but still, we need a lot of experience to start investing in stocks confidently. Many factors like political news, social media buzz and news related to current development in the company can create immediate change in stocks price movement. If we tie together such details like using technical and fundament analysis to understand the strength of stocks and sentimental analysis to understand the future direction of stocks. This can help us in predicting the stock price movement director with better accuracy.

In this project, we research techniques that we can use to make investment decisions better. Good prediction models will offer a better understanding of stock price movement and associated risks. A good decision can impact their returns on equity.

Issues with Traditional Methods

Investing in stocks provides satisfactory returns if money is put in and withdrawn in an intelligent manner, but it involves a lot of risks. The Stock market is a very appealing and interesting area where it can reward you with a good return if strategies implemented works the way you wanted, at the same time if it doesn't work as expected it can cost you a lot of fortune.

There have been numerous attempts made to invent methods that can accurately predict stock price movement but no methods succeeded in providing 100% accurate results. There are many factors (i.e. Company performance, Industry Performance, Investor sentiments, and Economy – Interest rates, Economic outlook, Inflation, Deflation, Economic and political shocks, economic policy change, the value of currency) that affect stocks price and complexity of these attributes make it difficult to predict the market correctly. These attributes generate lots of data daily. Handling this huge amount of data increases the complexity of the task which is not easy to do through traditional methods. Also, stock markets are very dynamic and change in stock price happens within seconds. Traditional methods are not efficient at making decisions because they require a lot of time to process the information before they can make a decision.

How Data Science can help?

Making profits in the stock market requires a lot of market knowledge and we need to learn how market behavior is based on past information to make the right judgment. Hence if we can build a 'historic-data-driven' predictive model that can predict probable market trends that such a model can be tremendously helpful for potential investors and would save them from the annoyance of analyzing data manually. This can be done using data science techniques. Data science

techniques are very efficient at analyzing past data and retrieving insight from the data dumps.

Machine learning techniques can be used to analyze historical data related to stocks, economy, policy change, and industry to help understand how stocks have been performing in the past. This data can be used to predict how it's going to perform in the future. Also, this data can be tied with other information like real-time news to make the prediction even more accurate.

Online social networks like Twitter and Facebook are permitting people who are zealous about trading and investing to break critical financial news faster and they also go deeper into relevant areas of research and sources leading to real-time insights. We can use NLP (natural language processing) and different machine learning techniques to analyze sentiments and market news to predict the movement of stocks.

We can take the example of Eurekahedge hedge funds. The firm Eurekahedge has published some interesting data. The graph below displays the performance of the Eurekahedge AI/Machine Learning Hedge Fund Index vs. traditional quant and hedge funds from 2010 to 2016 and we can see AI beats traditional funds.

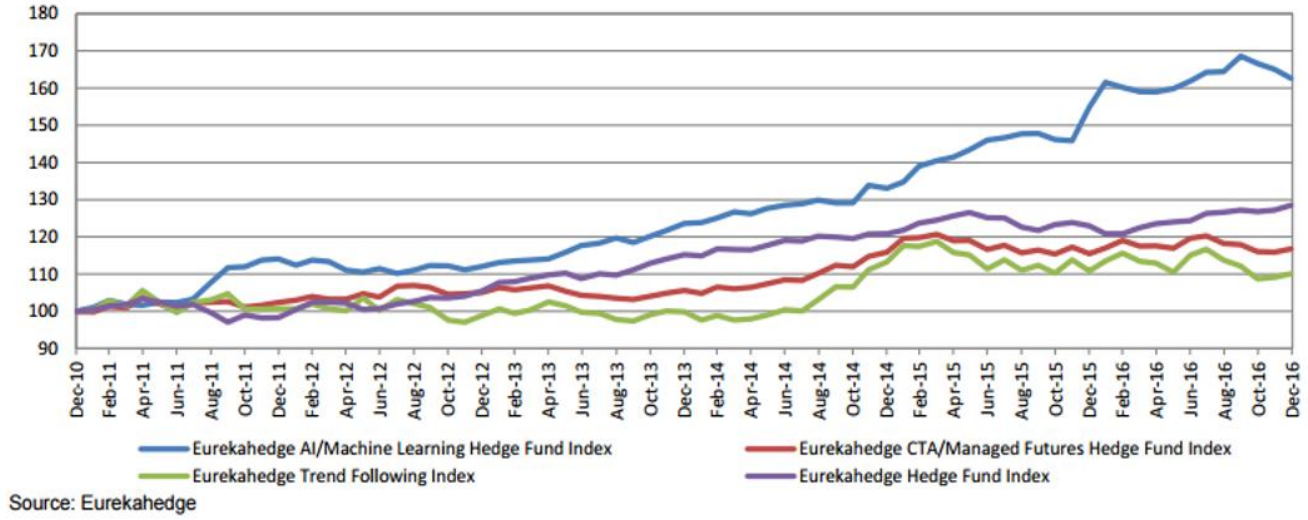

Figure 2: AI/Machine Learning Hedge Fund Index vs. quants and traditional hedge funds.

Data Science Techniques for Stock Prediction

We can use machine learning models and ingest sentiment data from trending financial news and social media buzz to predict stock price movements. There are various techniques that can be utilized to understand market sentiments and make predictions accordingly. Natural language processing can be utilized to understand language data. Classification & regression methods or artificial neural networks can be used to make the prediction. Below are a few of the techniques that are majorly used in Stock Market prediction:

Sentimental Analysis

Sentimental analysis can be applied to financial news headlines and social media buzz to understand the sentiments of people related to stocks. The best approach for sentimental analysis is the Lexicon Based Approach. This approach is based on the assumption that the sum of the sentimental orientation of each word can provide contextual sentimental orientation.

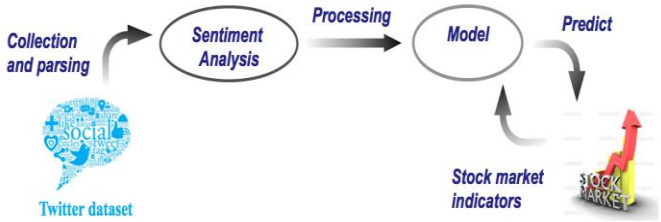

Figure 3: The picture shows an example of how to use Social network data to predict stock price movement using data from Twitter. Picture created by Iulian Vlad Serban, David Sierra Gonzalez, and Xuyang Wu for "Prediction of changes in the stock market using Twitter and Sentiment Analysis".

Support Vector Regression (SVR)

SVR is similar to Support Vector Machines (SVM) classifier and works on the same principle but with few minor differences. It tried to find optimal regression function with maximum margin (maximum flatness in data points) and minimal errors.

Artificial Neural Networks (ANN)

ANN models are one of the most popular prediction models. The ANN model can perform any function of the desired accuracy with just one hidden layer when used sufficient complexity in the model.

Back Propagation Neural Networks (BP-Neural Nets)

BP models are one of the most popular prediction models for financial forecasting. The BP network has been widely used in the area of financial time series forecasting because of its broad applicability to many business problems and its preeminent learning ability.

Hybrid Models

Hybrid models are used to predict stock prices. Hybrid models provide better accuracy than conventional models. For example, some hybrid model of genetic algorithm (GA) uses ANN with BP algorithm to train the ANN model by BP algorithms to optimize weights and bias values. This increases the accuracy of the ANN model. Similarly, the combined SVM-ICA model can be used to increase the prediction capability of the SVM classifier. The Independent component analysis (ICA) model can be used to select important variables from technical and fundamental analysis and feed selected data to the SVM model. Filtered data increases the prediction capability of the model.

Conclusion

Based on the study done through this project I came to the conclusion that it is difficult to predict stock price with 100% accuracy because of the complexity of the nature of attributes affecting stocks and the dynamic nature of the stock market. Data Science techniques can help in predicting the direction of stock price movement based on historical data and the latest market news. There is more work required to utilize daily market data, global economic data, and various other data related to stocks to make the stock prediction more accurate. A lot of investment is being done by investment companies like Goldman Sachs, Equbot, etc. towards stock market prediction and related fields for innovation using Artificial Intelligence.

A lot of new companies are coming up with a concept of Robo-Advisor for investment in Funds. The application of these kinds of advisors is increasing day by day.

Acknowledgements

I thank Professor Mr. Kenneth Birdsong for providing continuous constructive feedback and peers for their valuable inputs and suggestions. I also thank all the authors of the reference papers and articles.

References

1. Leekha, A., Wadhwa, A., Jain, N., & Wadhwa, M. (2018). Understanding the Impact of News on Stock Market Trends Using Natural Language Processing and Machine. International Journal of Knowledge Based Computer Systems, 6(2), 23–30. Retrieved from http://ezproxy.bellevue.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=iih&AN=134635931&site=eds-live
2. Qiu, M., & Song, Y. (2016). Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model. PLoS ONE, 11(5), 1–11. Retrieved from https://doi-org.ezproxy.bellevue.edu/10.1371/journal.pone.0155133
3. SCHUMAKER, R. P., & HSINCHUN CHEN. (2009). Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZFinText System. ACM Transactions on Information Systems, 27(2), 12:1-12:19. Retrieved from https://doi-org.ezproxy.bellevue.edu/10.1145/1462198.1462204
4. Ihlayyel, H. A. K., Sharef, N. M., Nazri, M. Z. A., & bakar, A. A. (2018). An enhanced feature representation based on linear regression model for stock market prediction. Intelligent Data Analysis, 22(1), 45–76. Retrieved from https://doi-org.ezproxy.bellevue.edu/10.3233/IDA-163316
5. COCIANU, C.-L., & GRIGORYAN, H. (2016). Machine Learning Techniques for Stock Market Prediction. A Case Study of Omv Petrom. Economic Computation & Economic Cybernetics Studies & Research, 50(3), 63–82. Retrieved from http://ezproxy.bellevue.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=118643772&site=eds-live
6. Chaudhuri, T. D., Ghosh, I., & Eram, S. (2016). Application of Unsupervised Feature Selection, Machine Learning and Evolutionary Algorithm in Predicting Stock Returns: A Study of Indian Firms. IUP Journal of Financial Risk Management, 13(3), 20–46. Retrieved from http://ezproxy.bellevue.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=120027252&site=eds-live
7. DAS, D., SADIQ, A. S., AHMAD, N. B., & LLORET, J. (2017). Stock Market Prediction with Big Data Through Hybridization of Data Mining and Optimized Neural Network Techniques. Journal of

Multiple-Valued Logic & Soft Computing, 29(1–2), 157–181. Retrieved from http://ezproxy.bellevue.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=iih&AN=123624695&site=eds-live

8. GRIGORYAN, H. (2016). A Stock Market Prediction Method Based on Support Vector Machines (SVM) and Independent Component Analysis (ICA). Database Systems Journal, 7(1), 12–21. Retrieved from http://ezproxy.bellevue.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=iih&AN=118278115&site=eds-live
9. Alostad, H., & Davulcu, H. (2017). Directional prediction of stock prices using breaking news on Twitter. Web Intelligence (2405-6456), 15(1), 1–17. Retrieved from https://doi-org.ezproxy.bellevue.edu/10.3233/WEB-170349
10. Tüfekci, P. (2016). Classification-based prediction models for stock price index movement. Intelligent Data Analysis, 20(2), 357–376. Retrieved from https://doi-org.ezproxy.bellevue.edu/10.3233/IDA-160809
11. Reed, M. (2016). A Study of Social Network Effects on the Stock Market. Journal of Behavioral Finance, 17(4), 342–351. Retrieved from https://doi-org.ezproxy.bellevue.edu/10.1080/15427560.2016.1238371
12. Hájek, P. (2018). Combining bag-of-words and sentiment features of annual reports to predict abnormal stock returns. Neural Computing & Applications, 29(7), 343–358. Retrieved from https://doi-org.ezproxy.bellevue.edu/10.1007/s00521-017-3194-2
13. Sebastian Heinz Nov 9, 2017. A simple deep learning model for stock price prediction using TensorFlow. Retrieved from https://medium.com/mlreview/a-simple-deep-learning-model-for-stock-price-prediction-using-tensorflow-30505541d877
14. Vivek Palaniappan Oct 31, 2018. Using Machine Learning to Predict Stock Prices. Retrieved from https://medium.com/analytics-vidhya/using-machine-learning-to-predict-stock-prices-c4d0b23b029a
15. AISHWARYA SINGH, Oct 25, 2018. Stock Prices Prediction Using Machine Learning and Deep Learning Techniques (with Python codes). Retrieved from https://www.analyticsvidhya.com/blog/2018/10/predicting-stock-price-machine-learningnd-deep-learning-techniques-python/
16. Team Auquan Nov 1, 2017. Application of Machine Learning Techniques to Trading. Retrieved from https://medium.com/auquan/https-medium-com-auquan-machine-learning-techniques-trading-b7120cee4f05
17. Vivek Palaniappan Sep 23, 2018. Neural Networks to Predict the Market. Retrieved from https://towardsdatascience.com/neural-networks-to-predict-the-market-c4861b649371
18. Machine Learning for Trading – Topic Overview. Retrieved from https://sigmoidal.io/machine-learning-for-trading/
19. Factors that can affect stock prices. Retrieved from https://www.getsmarteraboutmoney.ca/invest/investment-products/stocks/factors-that-can-affect-stock-prices/
20. Serban, I. V., González, D. S., & Wu, X. Prediction of changes in the stock market using twitter and sentiment analysis.